

MULTI-MODAL INTERNET INTERFACE DEVICE AND METHOD THEREOF

Publication number: KR20010076789 (A)
Publication date: 2001-08-16
Inventor(s): HWANG GYU UNG [KR]; KIM HOE RIN [KR]; PARK JUN [KR]
Applicant(s): KOREA ELECTRONICS TELECOMM [KR]
Classification:
- **international:** **G06F3/16; G06F3/16;** (IPC1-7): G06F3/16
- **European:**
Application number: KR20000004144 20000128
Priority number(s): KR20000004144 20000128

Abstract of **KR 20010076789 (A)**

PURPOSE: A multi-modal Internet interface device and a method thereof are provided to dialogue to a computer by a voice and to perform a web browsing by appending a voice interface function in a currently used web browser. **CONSTITUTION:** An MXML editor(312) makes a user prepare an MXML document using a template library by operations(311). A multi-modal Internet server(310) stores a prepared MXML document and includes an MXML document server providing the MXML document in accordance with the user. An MXML browser(321) is connected to the multi-modal Internet server (310) through the Internet, and interprets the MXML document and showing the contents of the MXML document on a screen and outputs a voice compounding message and prepares a voice recognition engine. A voice/compounding engine(322) performs voice recognition or a voice compounding. A TAPI (telephony application programming interface)(325)/MTAPI(multimedia telephony application programming interface)(326) provides an interface in an access to a multi-modal Internet client using a telephone(330). An I/O(input/output) device(324) comprises a keyboard, a mouse, a monitor, a mike, and a speaker. An I/O interface(323) provides an interface among the TAPI(325), the MTAPI(326), and the MXML browser(321).

.....
Data supplied from the **esp@cenet** database — Worldwide

(19) 대한민국특허청(KR) (12) 공개특허공보(A)

(51) 。 Int. Cl. 7
G06F 3/16

(11) 공개번호 특2001- 0076789
(43) 공개일자 2001년08월16일

(21) 출원번호 10- 2000- 0004144
(22) 출원일자 2000년01월28일

(71) 출원인 한국전자통신연구원
오길록
대전 유성구 가정동 161번지
(72) 발명자 황규웅
대전광역시서구월평3동다모아아파트111- 110
박준
대전광역시유성구전민동엑스포아파트505- 1202
김회린
대전광역시유성구전민동나래아파트105- 1001
(74) 대리인 전영일

심사청구 : 있음

(54) 멀티모달 인터넷 인터페이스 장치 및 방법

요약

본 발명은 컴퓨터를 사용하여 인터넷을 사용할 때 기존의 마우스, 키보드, 모니터 등을 이용하는 방식에 음성을 추가하여 인터넷을 이용할 수 있게 하는 방법 및 장치에 관한 것으로, HTML(HyperText Markup Language) 문서에 대화 모델을 추가하는 방법과 대화 모델을 쉽게 표현할 수 있는 메타 문법 및 이러한 서비스를 제공하는 멀티모달 인터넷 인터페이스 장치 및 방법을 제공한다.

본 발명에 따른 장치 및 방법으로 HTML로 기술된 웹 문서에 대화 모델을 작성하여 추가하게 되면 웹 브라우저를 사용하는 사용자는 화면에 표시되는 웹 페이지를 보고 마우스, 키보드를 사용하는 것에 추가하여 합성되는 음성을 듣고 자연스러운 음성으로 선택 및 질문을 하여 대화를 진행하면서 웹을 사용할 수 있게 된다.

대표도
도 3

명세서

도면의 간단한 설명

도 1은 본 발명의 멀티모달 인터넷 인터페이스의 사용예를 나타낸 도면

도 2는 MXML의 개념도

도 3은 본 발명의 멀티모달 인터넷 인터페이스 시스템의 서비스 구성도

* 도면의 주요부분에 대한 부호의 설명 *

310 : 멀티모달 인터넷 서버 311 : 작업별 템플릿 라이브러리

312 : MXML 문서 서버 320 : 멀티모달 인터넷 클라이언트

321 : MXML 브라우저 322 : 음성인식/합성 엔진

323 : I/O 인터페이스 324 : I/O 장치

325 : TAPI 326 : MTAPI

330 : 전화 331 : 음성전용 전화

332 : 멀티미디어 전화

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 컴퓨터를 사용하여 인터넷을 사용할 때 기존의 마우스, 키보드, 모니터 등을 이용하는 방식에 음성을 추가하여 인터넷을 이용할 수 있게 하는 방법 및 장치에 대한 것이다.

인간과 컴퓨터간의 인터페이스를 위해서 현재는 키보드, 마우스 등을 주로 사용하고 있지만 인간에게 보다 자연스러운 방법은 인간간에 이미 사용하고 있는 음성을 이용하는 방법이다.

이러한 방법은 이미 시도되고 있지만 컴퓨터의 경우 사람간의 대화와는 다르게 입력된 음성의 내용을 이해하는 것이 아니고 발성된 내용을 그대로 문자로 바꾸어 그 내용이 키보드로 입력된 것과 같이 반응하고 있는 상태로 음성 이해의 기술은 아직 활용되고 있지 않다. 그 이유는 음성이해 기술은 인공지능 기술과 마찬가지로 컴퓨터의 지능을 요구하며 현재 기술수준으로는 일반적인 분야에서는 불가능하고 여행계획 등과 같이 극히 제한된 분야에서만 가능하다.

음성을 이용하여 컴퓨터와 대화하는 기술이 많이 사용되는 곳은 ARS 시스템이다. 현재는 주로 음성합성을 통해 컴퓨터가 사람에게 메시지를 주고 사람은 전화기의 숫자판을 이용하여 원하는 의도를 전달하고 있다. 최근에는 단어 수준 및 간단한 문장 형태의 음성인식 기술을 통합하여 사람도 숫자판 대신에 음성을 이용하여 의도를 전달하는 시스템들도 개발되어 어느 정도 음성을 이용한 대화의 형태를 갖추고 있다.

그러나 ARS 시스템의 기본적인 문제는 원하는 정보나 서비스를 얻기 위해서는 복잡한 형태의 메뉴 트리를 따라 가야 한다는 것이다. 관공서의 ARS 시스템이나 기차예약을 해 본 사람들은 알겠지만 원하는 정보나 서비스를 얻기 위해서 수 많은 버튼을 누르고 특히 기차역의 코드를 알지 못하는 경우에는 안내 메시지를 전부 들어야 선택을 할 수 있게 된다. 음성인식을 이용하는 경우에는 기차역의 코드 등의 문제는 해결되지만 여전히 복잡한 메뉴 트리를 거쳐야 한다는 문제는 남는다.

ARS 시스템의 또 다른 문제는 유지 보수가 어렵고 호환성의 문제도 있다는 것이다. 이러한 점을 해결하기 위해 ARS 기반 시스템과 ARS에 사용되는 대화 모델 즉 메뉴 트리를 분리하는 VXML(Voice eXtensible Markup Language)을 IBM, AT&T, Motorola, Lucent 등이 개발 중이다. VXML은 HTML과 유사한 형태의 마크업 언어를 이용하여 메뉴 트리를 작성하면 독립된 VXML 서버에서 VXML 문서를 서비스 하는 형태이다. 마치 HTML 문서를 만들어 웹 호스팅 업체에 보내 놓으면 웹 서버 등의 서비스는 웹 호스팅 업체에서 수행하는 것과 같다.

이와 관련된 또 다른 시도로 마이크로소프트사에서 제안한 발명이 있는데 그 내용은 ARS 시스템을 사용할 때 전화 만을 쓰는 것이 아니라 컴퓨터도 동시에 연결하여 ARS 메뉴 내용을 화면에 보여 주고 전화 숫자판 또는 컴퓨터의 마우스, 키보드 등을 사용하여 선택을 할 수 있게 하는 것이다. 이 때 ARS 메뉴의 전달은 HTML과 유사한 방식을 사용한다. 마이크로소프트사의 발명의 경우 기존 ARS 시스템의 단점을 해결하기 위해 컴퓨터 등의 화면에 메뉴 트리의 내용을 나타내어 메뉴 트리 따라가기를 쉽게 하였다. 하지만, 이 발명 역시 사용자가 컴퓨터의 마우스, 키보드 등을 사용하여 선택하여야 하며 음성으로는 선택이 불가능하다.

발명이 이루고자 하는 기술적 과제

상기한 종래 기술의 문제점을 해결하기 위한 본 발명은 기존의 웹 브라우징에서 음성인터페이스 기능을 추가하여 컴퓨터와 음성으로 대화하며 웹 브라우징을 수행할 수 있도록 하는 멀티모달 인터넷 인터페이스 장치 및 방법을 제공하는 데 그 목적이 있다.

이를 위하여 본 발명에서는 기존의 HTML 문서에 대화 모델을 추가하는 방법과 그 서비스 방식 및 컴퓨터와의 대화에서 음성인식을 위한 언어 모델로 메타 문법을 사용하는 것을 특징으로 한다.

발명의 구성 및 작용

이하, 첨부도면을 참조하여 본 발명을 상세하게 설명한다.

먼저 기존의 웹 브라우징에서 음성을 추가로 사용할 수 있는 시스템에 대해 기술한다. 시스템의 사용 예는 도 1에 나타내었다.

이와 같이 음성을 사용할 수 있는 시스템을 구성하기 위해 기존의 HTML 문서에 음성 관련 부분을 추가한 멀티모달 X ML(Multi-modal eXtensible Markup Language: 이하 MXML라 함) 문서를 제안한다.

MXML 문서는 하나의 문서에 HTML의 내용과 메타 태그(Meta tag)를 이용하여 음성 XML의 내용을 함께 기술한 것이다. 음성을 처리할 수 있는 브라우저에서는 음성 XML의 내용을 바탕으로 음성 인터페이스를 지원하게 되고 그렇지 못한 경우에는 음성 관련 내용은 무시하게 된다.

이를 도 2의 MXML 문서 개념도를 참조하여 좀 더 자세하게 설명한다.

도 2는 기존의 HTML과 음성 XML 및 대화관리가 어떻게 결합되어 MXML을 구성하는지 설명하고 있다. 기존의 HTML의 경우는 마우스와 키보드를 입력수단으로 사용하고 출력으로는 화면에 텍스트, 영상, 그리고 스피커를 통해 음성과 음향을 출력하고 있다. 이 경우 음성의 출력은 음성합성을 통한 것은 아니고 주로 녹음된 음성을 음향의 경우와 같은 방식으로 출력하게 된다. 기존에 제안된 음성 XML은 전화를 이용하는 경우를 기본적으로 가정하며 전화를 통해 음성을 입력하고 이를 음성 인식하여 처리하게 된다. 또 전화기의 키패드를 이용해 DTMF 신호도 입력하여 사용한다. 출력으

로는 합성 음성이 유일하다. 이 때 사용자는 전화를 통해 원하는 정보를 얻거나 예약 등의 작업을 하게 되므로 이를 처리하는 컴퓨터와의 대화관리가 필요하다. 도 2에서 설명하는 MXML은 이러한 기존의 두 가지 방법을 통합하여 음성 입출력이 가능하고 선택적으로 마우스, 키보드, 화면이 제공되는 단말기에서, 제공되는 다양한 입출력 수단을 함께 사용하여 보다 사용하기 편하고 효율적인 인터페이스를 제공하고자 한다.

도 3은 본 발명에 따른 멀티모달 인터넷 인터페이스 시스템의 구성도 이다.

도면에서 보듯이, 작업별 템플릿 라이브러리(311), 상기 작업별 템플릿 라이브러리(311)를 이용하여 사용자가 MXML 문서를 작성할 수 있도록 하는 MXML 편집기(312), 및 작성된 MXML 문서를 저장하고 사용자의 요청에 따라 MXML 문서를 제공하는 MXML 문서 서버를 포함하는 멀티모달 인터넷 서버(310); 상기 멀티모달 인터넷 서버(310)와 인터넷으로 연결되며, MXML 문서를 해석하여 HTML 내용은 화면에 보여 주고 음성 XML내용을 해석하여 음성 합성할 메시지는 음성 합성 엔진을 이용하여 합성하여 스피커 또는 전화 인터페이스를 통해 들려 주고, 음성 인식을 위한 언어모델을 이용하여 음성인식 엔진을 준비하고 사용자가 발성한 음성의 내용을 인식하여 음성 XML에 지정된 동작을 수행하는 MXML 브라우저(321)와, 상기 MXML 브라우저(321)의 요구에 따라 음성인식 또는 음성합성을 수행하는 음성 인식/합성 엔진(322)과, 전화(330)를 이용하여 멀티모달 인터넷 클라이언트에 접근하는 경우에 인터페이스를 제공하는 TAPI(Telephony Application Programming Interface : 325)/MTAPI(Multimedia Telephony Application Programming Interface : 326)와, 키보드, 마우스, 모니터, 마이크, 스피커 등의 I/O(Input/Output) 장치(324), 상기 TAPI(325) 및 MTAPI(326)과 상기 MXML 브라우저(321)와의 인터페이스를 제공하는 I/O 인터페이스(323)를 포함하는 멀티모달 인터넷 클라이언트(320)로 구성된다.

멀티모달 인터넷 서버(310)는 기존의 웹 서버의 역할을 하는 것으로 MXML 문서를 요청에 따라 제공해 주는 역할을 한다. 멀티모달 인터넷 서버(310)에는 MXML 문서가 저장되는데 이는 MXML 편집기(312)와 작업별 템플릿 라이브러리(311)를 이용해 작성하게 된다. MXML의 경우 일반 HTML과 달리 음성 인식을 위한 언어모델을 작성해 주어야 하는데 이는 언어 처리에 대한 전문적인 지식을 요구하므로 일반 HTML 작성자가 작성하기 어렵다. 이러한 문제를 해결하기 위해 사용자가 발성할 문장 그 자체를 모두 기술하는 일반 문법이 아닌 한 단계 위의 메타 문법과 작업별 템플릿 라이브러리를 사용하게 되는데 이 문법에 관해서는 후술한다.

멀티모달 인터넷 클라이언트(320)는 기존의 PC/WS(Personal Computer/Work Station)와 같이 CPU를 갖춘 단말기이다. MXML을 해석해서 화면에 보여 주는 MXML 브라우저(321)와 음성 인식/합성 엔진(322), I/O 장치(키보드, 마우스, 모니터, 마이크, 스피커 등 : 324), TAPI 및 MTAP와 MXML 브라우저(321)와의 인터페이스 모듈, 즉 I/O 인터페이스(323)를 가지고 있다.

MXML 브라우저(321)은 요청된 MXML 문서를 인터넷을 통해 멀티모달 인터넷 서버(310)에서 제공받아, 이를 해석하여 HTML 내용은 기존의 웹 브라우저처럼 화면(PC/WS의 모니터 또는 스크린 폰/인터넷 접속 휴대폰의 화면)에 보여준다. 그리고 음성 XML내용을 해석하여 음성 합성할 메시지는 음성 합성 엔진을 이용하여 합성을 해 스피커, TAPI(325) 또는 MTAPI(326)를 통해 들려 주고, 음성 인식을 위한 언어모델을 이용하여 음성인식 엔진을 준비하고 사용자가 발성한 음성의 내용을 인식하여 음성 XML에 지정된 동작을 수행한다.

TAPI(325)과 MTAPI(326)는 제3의 장소에서 전화(330)와 멀티모달 인터넷 클라이언트(320)를 이용하여 인터넷 브라우징을 할 수도록 한다.

전화(330)로는 사용되는 단말기로 음성만 사용할 수 있는 음성전용전화(331)와 스크린을 가지고 있어 화면 표시도 가능한 스크린 폰, 인터넷 접속 휴대폰 등의 멀티미디어전화(332)로 나눌 수 있다. 멀티미디어전화기(332)에 대해서는 음성뿐만 아니라 HTML의 내용도 같이 보내 준다.

I/O 장치(324)에서 키보드와 마우스, 모니터는 필수적인 것은 아니다. 이동형 단말기인 경우나 시스템의 구성상 뺄 수도 있다. 또 장애자의 경우에는 장애 종류와 정도에 따라 키보드, 마우스, 모니터 등을 사용하지 못할 수도 있다. 이러한 경우에도 스피커와 마이크를 이용하여 원하는 작업을 어느 정도 수행할 수 있다.

음성 XML의 내용에 대해 설명한다.

음성 XML의 목적은 컴퓨터와 사용자가 인간에게 자유로운 음성을 이용하여 대화를 나누어 사용자가 원하는 동작을 컴퓨터로 하여금 수행하도록 하는 것이다. 이를 위해서 음성 XML에는 사용자와 컴퓨터가 나눌 수 있는 대화의 내용을 기술하게 되는데 본 발명에서는 이를 메타 문법을 사용하여 기술하도록 한다. 음성 XML의 내용으로는 컴퓨터에서 사용자에게 제공하는 말과 사용자가 발성할 내용을 기술한 메타 문법, 그리고 메타 문법을 통해 사용자의 의도를 이해하고 이에 따른 동작의 지정이 있다.

메타 문법이라 함은 기존의 언어모델의 문법보다 한 단계 상위의 문법이라는 의미이다. 메타 문법은 두 가지로 나뉘는데 작업 지향적인 작업 메타 문법과 대화의 목적 지향적인 자료 메타 문법이 있다.

작업별 템플릿 라이브러리는 호텔예약, 기차표예매와 같이 그 자체로 하나의 작업이 완료될 수 있는 형태이다. 즉 일반적으로 많이 사용되는 작업에 대한 문법을 기술하여 놓은 것이다. MXML 작성자는 음성 XML의 내용에 해당되는 작업별 템플릿 라이브러리 파일의 위치를 지정하는 것만으로 컴퓨터가 사용자와 대화하여 사용자가 원하는 작업을 완료할 수 있도록 한다. 작업별 템플릿 라이브러리를 지정할 때 파라미터를 주어 철도 회사 명이나 호텔 명 등의 가변적인 정보를 바꿀 수 있다. 작업 지향적인 문법은 널리 여러 곳에서 공통적으로 사용되는 경우에 라이브러리의 형태로 문서 작성자가 아닌 제 삼자가 만들어 공급할 수 있을 것이다.

이에 반하여 자료 메타 문법은 대화의 좀 더 작은 부분에 적용된다. 그 예로는 사용자의 이름을 알아내는 대화나 신용카드 번호, 날짜 등을 알아내는 대화가 해당된다. 이 문법의 사용 예는 다음과 같다.

예)

```
get_data(costumer_name)
```

```
get_data(credit_card_number)
```

기존의 문법에서는 사용자의 이름을 알아 내는 경우에 사용자가 이름만을 말하게 하도록 제한하거나 몇 개의 표현만을 지원하고 있지만 이 메타 문법의 경우에는 사용자와 해당되는 주제를 놓고 대화를 나누어 원하는 정보를 얻어 내도록 동작한다. 사용자가 해당되는 주제에 대해 발성할 수는 대표적인 대부분의 표현을 처리할 수 있도록 한다.

발명의 효과

상술한 바와 같은 본 발명은 기존의 인터넷 브라우저에서 키보드, 마우스, 모니터 등의 인터페이스 방식 이외에 사람에게 편리한 음성을 인터페이스로 추가함으로써 보다 편리하게 인터넷 브라우저를 할 수 있다. 이 때 대화의 방식은 기존의 고정된 표현만을 사용하는 방식이 아니라 메타 문법 기능을 통해 구현되는 일상생활에서 사용하는 자유로운 형태의 대화 방식을 채용하여 사용자에게 편의를 제공한다.

이상에서 본 발명에 대한 기술사상을 첨부도면과 함께 서술하였지만 이는 본 발명의 바람직한 실시예를 예시적으로 설명한 것이지 본 발명을 한정하는 것은 아니다. 또한, 이 기술분야의 통상의 지식을 가진 자라면 누구나 본 발명의 기술 사상의 범주를 이탈하지 않는 범위 내에서 다양한 변형 및 모방이 가능함은 명백한 사실이다.

(57) 청구의 범위

청구항 1.

기존의 마우스, 키보드, 모니터 등을 이용하는 것 외에 음성을 통해서도 인터넷 서비스가 가능하도록 하는 멀티모달 인터넷 인터페이스 장치에 있어서,

작업별 템플릿 라이브러리와, 상기 작업별 템플릿 라이브러리를 이용하여 사용자가 MXML 문서를 작성할 수 있도록 하는 MXML 편집기와, 및 작성된 MXML 문서를 저장하고 사용자의 요청에 따라 MXML 문서를 제공하는 MXML 문서 서버를 포함하는 멀티모달 인터넷 서버;

상기 멀티모달 인터넷 서버와 인터넷으로 연결되며, MXML 문서를 해석하여 HTML 내용은 화면에 보여 주고 음성 XML내용을 해석하여 음성 합성할 메시지는 음성 합성 엔진을 이용하여 합성하여 스피커 또는 전화 인터페이스를 통해 들려 주고, 음성 인식을 위한 언어모델을 이용하여 음성인식 엔진을 준비하고 사용자가 발성한 음성의 내용을 인식하여 음성 XML에 지정된 동작을 수행하는 MXML 브라우저와, 상기 MXML 브라우저의 요구에 따라 음성인식 또는 음성합성을 수행하는 음성 인식/합성 엔진과, 전화를 이용하여 멀티모달 인터넷 클라이언트에 접근하는 경우에 인터페이스를 제공하는 TAPI 및 MTAPI와, 마이크, 스피커 등을 포함하는 I/O 장치와, 상기 TAPI 및 MTAPI과 상기 MXML 브라우저와의 인터페이스를 제공하는 I/O 인터페이스를 포함하는 멀티모달 인터넷 클라이언트로 구성되는 것을 특징으로 하는 멀티모달 인터넷 인터페이스 장치.

청구항 2.

기존의 마우스, 키보드, 모니터 등을 이용하는 것 외에 음성을 통해서도 인터넷 서비스가 가능하도록,

HTML 문서에 음성 XML 부분을 추가하여 MXML 문서를 작성하고,

음성 XML의 내용으로 컴퓨터에서 사용자에게 제공하는 말과 사용자가 발성할 내용을 기술한 메타 문법, 및 메타 문법을 통해 사용자의 의도를 이해하고 이에 따른 동작의 지정을 포함하여 음성 및 화면을 통해 인터넷 서비스가 가능하도록 하는 것을 특징으로 하는 멀티모달 인터넷 인터페이스 방법.

청구항 3.

제2항에 있어서,

음성 XML의 내용은 그 자체로 하나의 작업이 완료될 수 있는 형태인 경우 작업 지향적인 작업 메타 문법을 사용하여 기술하고,

사용자와 해당되는 주제를 놓고 대화를 나누어 원하는 정보를 얻어낼 필요가 있는 부분은 사용자가 해당되는 주제에 대해 발성할 수 있는 대표적인 대부분의 표현을 처리할 수 있도록 목적 지향적인 자료 메타 문법을 이용하여 기술하는 것을 특징으로 하는 멀티모달 인터넷 인터페이스 방법.

도면

도면 1

안녕하세요. 롯데 호텔입니다. 무엇을 도와드릴까요?

손님께서 예약하시고자 하는 지점을 선택해 주십시오.

원하시는 객실을 선택해 주세요.

언제 숙박하실 예정입니까?

손님의 성함과 신용카드 번호를 입력하거나 말씀해 주십시오.

예약되었습니다. 이용해주셔서 감사합니다.

방을 예약하려고 하는데요.

대전이요.

스탠더드룸으로 주세요.

십일월 이십삼일이요.

이름란에 "홍길동", 카드번호란에 7120 2324 5568 8184 입력

Reservation

Check Availability

Reservation

| ROOM | TYPE | PRICE | SELECT |
|--------------------|------|----------|--------------------------|
| Standard Room | | ₩110,000 | <input type="checkbox"/> |
| Deluxe Twin Room | | ₩110,000 | <input type="checkbox"/> |
| Family Twin Room | | ₩120,000 | <input type="checkbox"/> |
| Superior Onnd Room | | | <input type="checkbox"/> |
| Deluxe Onnd Room | | | <input type="checkbox"/> |
| Superior Suite | | | <input type="checkbox"/> |

Reservation

Property: HOTEL LOTTE TAEJON - TAEJON

Room Type: Standard ROOM Double

Check In: [Date] [Time]

Check Out: [Date] [Time]

Arrival Flight Number: [Optional]

Number of Rooms: [Required]

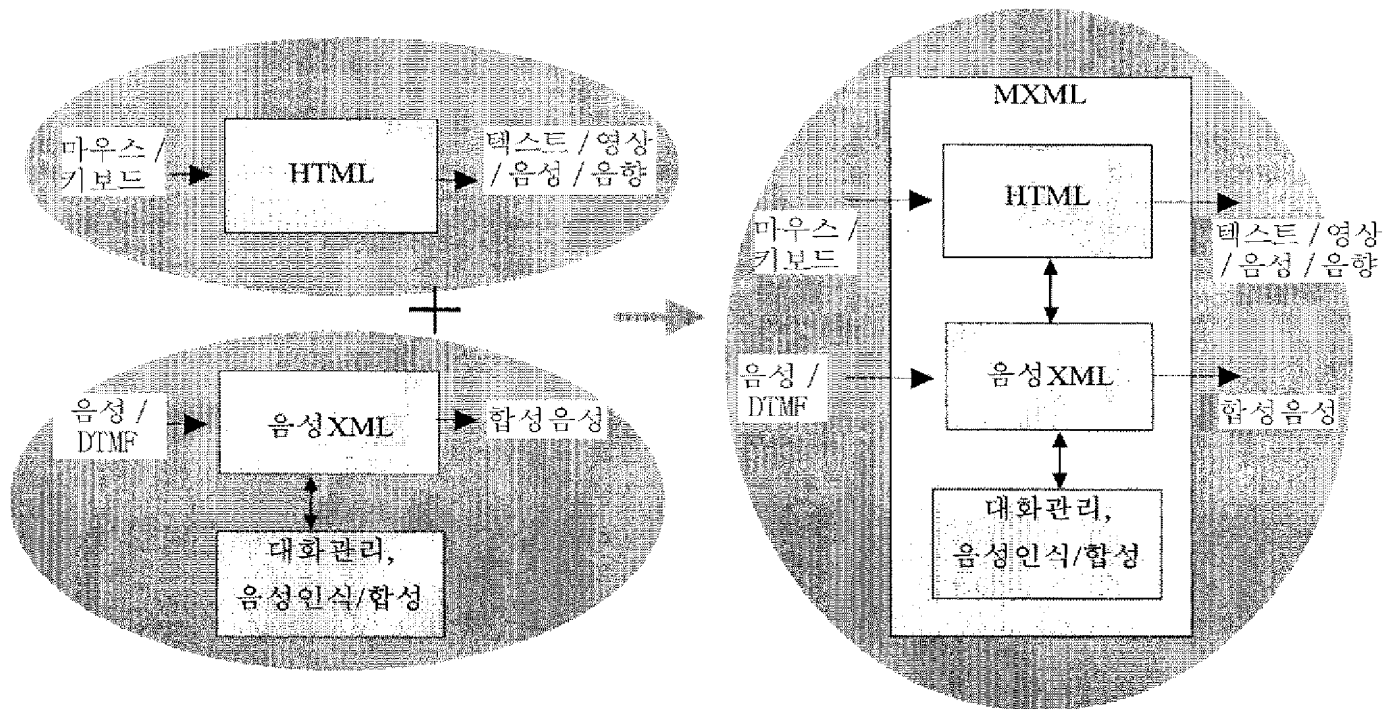
Number of Adults: [Required]

Number of Extrabed: [Optional]

Reservation

| ROOM | Standard Room |
|------|-----------------|
| 가격 | ₩110,000 |
| 기간 | 1000.11.23 ~ 24 |
| 요금 | ₩110,000 |

도면 2



도면 3

